Towards an Adaptive System: Users' Preferences and Responses to an Intelligent Virtual Advisor based on Individual Differences

Hedieh Ranjbartabar *hedieh.ranjbartabar@students.mq.edu.au*
Department of Computing, Macquarie University
Sydney, Australia

Deborah Richards *deborah.richards@mq.edu.au*
Department of Computing, Macquarie University
Sydney, Australia

Cat Kutay *cat.kutay@uts.edu.au*
Faculty of Engineering and IT,
University of Technology Sydney, Australia

Samuel Mascarenhas *samuel.mascarenhas@gaips.inesc-id.pt*
INESC-ID & Instituto Superior Técnico
Universidade de Lisboa

Abstract

Understanding individual differences in students could help Intelligent Virtual Agents (IVAs) to provide tailored educational and emotional support. Towards creating a student model that the agent can reason over and adapt to accordingly, we conducted a study to identify possible relationships and rules based on the students' personality, emotional state and character preferences. The purpose of our virtual advisors was to "Reduce Study Stress". The experiment with 73 participants, consisting of one within-subjects factor (virtual advisors with empathic and neutral dialogue) and one between-subjects factor (different order of receiving empathic and neutral advisors), formed two experimental and one control groups. We measured preferences, perceived helpfulness and study stress level. Groups using the IVAs reported significantly lower levels of study stress at the end of the study. Some differences were found in preferences for and responses to IVA behaviour based on participants' gender, personality and levels of depression, anxiety and stress.
Keywords: Intelligent Virtual Agents, User/Student Modelling, gender, personality, DASS21.

1. Introduction

Intelligent Virtual Agent (IVA) technology has reached a level of sophistication that allows IVAs to exhibit congruent and socially plausible behaviours through having situation awareness [1], verbal and non-verbal communication skills and their own memories [2], personality [2], cultural norms [3] and emotion appraisal systems [4]. This state of the art has been achieved through a focus on IVA believability and social capability. We believe that IVA technology has matured to a level where it is time to turn our attention from the agent to the human to identify what attributes, behaviours and capabilities does an IVA need to best serve the needs of a specific human in a specific context. A successful human teacher adapts their expectations and behaviours according to knowledge of their student. As software, it should be possible to adapt IVAs so that their appearance, memories and background are what is needed by the human to bring about belief and behaviour change and provide long term support.

In health and wellbeing applications, computer-mediated interactions via an IVA can have advantages over human interactions including increased accessibility, confidentiality and divulgation; tailored information; diminished variability; avoidance of righting reflex with

infinite patience; addressing low literacy, lower attrition rates; expressing empathy; allowing patient-physician concordance/ matching and provide working alliance [5]. Given the importance of the learner's emotional state in achieving learning outcomes [6], empathic agents (i.e. agents that respond in emotionally sensitive ways) may play important roles not only in health and wellbeing applications, but also in educational contexts. IVAs with educational roles are often called animated pedagogical agents (APAs) [7]. APAs have been shown to assist learning and elicit emotional reactions from learners [8] [9]. Due to the influence of the learner's emotion on their learning, there are many approaches to elicit emotions from students, e.g. observation [10], self-reporting [11], emoticons [12], measuring physiological data [13, 14] and detection of facial expressions [15, 16]. This research was motivated by earlier studies on adaptive agents. Our novel contribution extends that work by investigating learner preferences for an IVA, the impact of the IVA on the learners' emotions and how learners respond to the use of empathy by an IVA based on their individual differences. Towards understanding these issues, this paper seeks to answer the research questions:

1) *Do the users' gender, age, personality or psychological state influence their 1- preferences for a character; 2- responses to the IVAs and 3- derived benefit (i.e. reduced study stress) from interacting with the IVA?*
2) *Do these differences suggest variations in IVAs and how IVAs should adapt to their user?*

In the next section, we provide an overview of related background on adaptive agents. In Section 3 we describe our methodology followed by the results in Section 4. We end with discussion of our findings and limitations (Section 5) and conclusion and future work (Section 6).

2. Adaptive & Customisable Agents

Although there are number of studies on adaptive virtual environments (e.g. [17], [18]), in this paper we focus on the literature on adaptive virtual agents without considering the virtual environments. A number of different types of adaptive agents have been created. A key way of distinguishing these agents is according to the number of modalities they support. A mimicking agent [19] may just copy or mirror the non-verbal behaviours of the human, a listening agent [1, 20], may respond according to the inputs (verbal and non-verbal) of the human in an emotionally congruent way. A laughing agent [21] may tell a joke and laugh with the human, with limited understanding of whether or why the human is laughing and a culturally adaptive agent [22] may copy gestures and postures of the user. While existing research does model factors for agents such as personality [2] and culture [3] and allow the agent to respond in emotionally congruent ways [4], tailoring behaviours in real-time based on individual differences, such as personality or culture, is largely an open question.

IVA adaptation to the user could relate to the appearance of the character, which can represent a particular gender, age or ethnic background. Research has found that the appearance of a character can influence user perception (e.g. [23], [24]) and other work has explored the use of IVAs in helping contexts including fitness and healthier eating [25], reducing alcohol consumption [26] and stress management [27]. However, studies that combine the manipulation of appearance and its effect on the usefulness of the IVA are still lacking.

One approach to provide an IVA model that matches the user's preferences is to allow them to create their own avatar. This can results in participants focusing on their character and also creating characters of their idealized self Ducheneaut, Wen [28], Being assigned an avatar, however, has been found to influence how the player/user behaves based on features of the avatar such as players disclosing more when they perceive the avatar as more friendly due to a more attractive appearance. There have also been difference in perceptions of avatars by different age groups. For example, in a system to teach children how to handle bullying situations, Hall, Woods [29] children's scores were more positive than the teachers and male children in particular found the storyline more believable. In another study, virtual doctors were perceived to be more knowledgeable by females if the virtual human's body mass index (BMI) was discordant while males deemed the virtual human to be more knowledge if BMIs were

similar [30] Bailenson, Swinth [23] found that the self-reported sense of co-presence increased when the character better resembles a human in looks and behaviour.

The main current virtual agent architectures (i.e. FAtiMA[31], GRETA [32], Listening Agent [20], ODVIC [26] and SimSensei [1]) have been developed and used for different purposes. There is still no unified user model in the current agent architectures to interpret the inputs from the user and respond adaptively. As an example, FAtiMA integrates emotional appraisal theories (i.e. OCC [9a]) to have emotion and personality influence the behaviour of the character. Modelling of emotions relies on appraisal theories based on evaluations of events in the environment. Although in FAtiMA the agent has its own personality, beliefs and emotions, the agent is not capable of assessing the personality of the user or adapting according to individual differences in users. Our work towards developing a user model, specifically a student model in the context of educational applications, is motivated by prior work showing that individual features of the user, such as gender or background, influence their preferences and responses towards different agents and addressing the current gap in agents' ability to detect and adapt accordingly.

3. Methodology

In this study, we explored the influence of the users' personality, gender, age and psychological emotional state on preferences for an IVA to help them and their responses to two IVAs, one with neutral and the other with empathic dialogue. We wanted to see if certain individuals preferred one type of dialogue over the other. We designed an experiment consisting of one within-subjects factor (empathic and neutral virtual advisor) and one between-subjects factor (different order of the experimental sequence), forming two experimental groups and one control group (Figure 1) who received a pdf document to assist them.

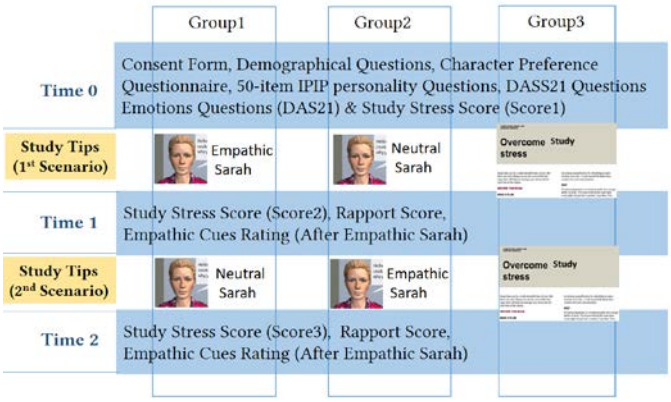

Fig. 1. Experimental/Control groups design

Participants were invited via the online recruitment program of the Psychology Department. All participants were volunteers and could choose our study half an hour course credit for their participation. The experiment was approved by the Macquarie University's Human Research Ethics Committee.

3.1. Materials and Methods

We created a meaningful scenario for our population concerning "Reducing Study Stress" involving two versions (empathic/neutral) of Sarah, a virtual advisor who provided tips to students for reducing their study stress.

Fig. 2. Sarah, the virtual advisor

After informed consent and logging into the system, Sarah (Figure 2) introduced herself and provided study tips in two rounds. We created two dialogues for the three study groups (scenario 1 and scenario 2), with two versions (empathic and neutral). The content of the dialogues was derived from the campus wellbeing and support service of our university, which included tips about work, study and life balance; exercise and healthy eating; overcoming exam stress and socialising. For consistency and comparison, these tips were structured into two documents and presented to the control group

in the same order as the dialogues. Sarah was developed based on the FAtiMA agent architecture and the Unity3D game engine. We chose FAtiMA [31] because it is a cognitive agent architecture for creating autonomous, engaging and believable characters. We plan to extend FAtiMA to include an explicit user model and rules and/or modules that will allow the IVA to make decisions regarding its behaviour that are tailored to that user.

To design Sarah's empathic dialogue we used the empathic cues identified from the literature [33]. Sarah's verbal behaviours seek to establish rapport with the user. The empathic cues used in her dialogue with related dialogue samples are shown in Table 1. Both versions of Sarah conveyed a pleasant smiling face and included lip-synching. No other non-verbal behaviours were used. Dialogue was spoken by Sarah using TextToSpeech generation of an Australian synthesized voice. Users responded through selection of answer options.

Table 1. Empathic Cues used in the Dialogues

Empathic Cues	Example
Social Dialogue	I hope you have something nice planned for your weekend.
Meta-Relational Dialogue	I'm hoping to give you some ideas that might help you with your study and life in general.
Empathic Feedback	Oh dear, Sounds like you have some stuff happening. I am here for you.
Humor	Life can box you in – look at me, I'm stuck inside this machine – just kidding.
Continuity behaviors	It's nice to see you again. I'm hoping I can get to know you better.
Self-Disclosure	I also get really stressed when I have lots of study.
Meta-Relational Dialogue & Sharing knowledge	We're doing some great work together.
Mirroring	I do the same.
Politeness	Please make yourself comfortable.
Inclusive pronoun	Sometimes it's nice to have our own time.

3.2. Procedure and Data Collection

The Qualtrics research software (Qualtrics.com) provided participants with online access to all the intervention materials and surveys. It was our goal to understand the human and how individual differences might influence preference for features and behaviours of the IVA. Thus, first we collected demographic data including gender, age, cultural background, degree being studied, computer game activity and attitude towards study (using 5pt Likert scale from strongly dislike to strongly like) followed by a preference questionnaire (see results).

We also wanted to identify if the humans' personality influenced their preferences and responses to the IVAs. Therefore we made use of International Personality Item Pool (IPIP) which is a personality inventory to model personality of the users [34]. The version used in this experiment contains 50 items, 10 items measuring each of the five personality scales. The scales are known as the Big-Five personality factors and include Openness to experience (Intellect/Imagination) (O), Conscientiousness (C), Extraversion (E), Agreeableness (A) and Neuroticism (conversely, Emotional Stability) (N). We also chose to use Depression-Anxiety and Stress (DASS21) questionnaire to capture the psychological emotional state of the user [35].

Next, participants were randomly allocated to one of the three designed groups using the Qualtrics Randomizer feature and they downloaded and ran the corresponding executable Unity3D file (groups 1 and 2) or downloaded the pdf files for the control group. For groups 1 and 2, the users' keyboard and mouse interaction data with the IVA were captured into a separate MySQL database.

To identify if the study tips were useful and whether the IVA was "fit for purpose", we also asked the students' "Study Stress Score" at three time points (before and after each scenario) by asking: think about your emotional feeling towards your study on a scale of 0 to 10. Zero means "extremely good and relaxed" and 10 means "extremely bad and stressed". To evaluate the empathic dialogue provided by Sarah, we chose one statement for each of the empathic cues

that the participant experienced and asked them to indicate if they found it empathic, helpful and/or stupid.

4. Results

To find if there are differences in participants' responses to the IVA according to different user profiles, our results include detailed user profiling including demographics (4.1), study stress score result, the changes in participants' study stress levels following conversations with Sarah (4.2) and the results of comparisons of individual differences with responses (4.3).

4.1. Participants

In total our study involved 73 university students from Australia, aged between 18-33 (mean age=20.09, SD=2.56). Only one participant was over 30 years old. Due to the homogeneity of the age of the population, we do not present any age specific results and did not conduct any analysis by age. Table 2 shows the gender distribution across each group.

Table 2. Gender Distribution Across Experimental Groups

Group	Female	Male	Other	Total
1- Empathic-Neutral	15	10	0	25
2- Neutral-Empathic	12	10	1	23
3- Control	12	13	0	25
Total	39	33	1	73

Different cultural groups were involved in the study. The largest group is Oceania (27.4%) and the next largest groups are Northern-Western European and South-East Asian (17.81% each). Most of the participants were enrolled in a Psychology degree (73.97%) and less than half of the participants regularly played computer games (45.21%). The majority of participants' were neutral about study (54.79%) while 19.18% and 4.11% of students' chose like and love, and 17.18% and 4.11% of students' chose dislike and hate, respectively. Table 3 shows categorised personality results.

Table 3. Personality Dimension Distribution

	Low		Medium		High	
Openness	2	2.74%	58	79.45%	13	17.81%
Conscientious	8	10.96%	58	79.45%	7	9.59%
Extravert	16	21.92%	46	63.01%	11	15.07%
Agreeable	1	1.37%	39	53.92%	33	45.21%
Emotionally Stable	20	27.4%	49	67.12%	4	5.48%

DASS21 results by gender are shown in Table 4. Table 5 shows frequency and percentages by gender for participants' character preferences before interaction.

Table 4. DASS 21 Results

		Female		Male		Other		Total	
		N	%	N	%	N	%	N	%
Depression	Normal	10	25.64%	11	33.33%	0	0	21	28.77%
	Mild	5	12.82%	7	21.21%	0	0	12	16.44%
	Moderate	13	33.33%	7	21.21%	0	0	20	27.40%
	Severe	4	10.26%	2	6.06%	1	100%	7	9.59%
	ExtremelySevere	7	17.95%	6	18.18%	0	0	13	17.81%

		Female		Male		Other		Total	
		N	%	N	%	N	%	N	%
Anxiety	Normal	10	25.64%	5	15.15%	0	0	15	20.55%
	Mild	5	12.82%	6	18.18%	0	0	11	15.07%
	Moderate	8	20.51%	5	15.15%	0	0	13	17.81%
	Severe	2	5.13%	4	12.12%	0	0	6	8.22%
	ExtremelySevere	14	35.90%	13	39.39%	1	100%	28	38.36%
Stress	Normal	23	58.97%	18	54.55%	0	0	41	56.16%
	Mild	5	12.82%	6	18.18%	0	0	11	15.07%
	Moderate	4	10.26%	2	6.06%	1	100%	7	9.59%
	Severe	5	12.82%	4	12.12%	0	0	9	12.33%
	ExtremelySevere	2	5.13%	3	9.09%	0	0	5	6.85%

Table 5. The Preference Questionnaire Result

Virtual Character Preference		Female		Male		Other		Total	
Would you prefer a virtual character to be?	younger than you	0	0%	4	12.12%	0	0%	4	5.48%
	older than you	9	23.08%	6	18.18%	0	0%	15	20.55%
	same age as you	30	76.92%	23	69.70%	1	100%	54	73.97%
	male	2	5.13%	14	42.42%	0	0%	16	21.92%
	female	19	48.72%	4	12.12%	0	0%	23	31.51%
	doesn't matter	18	46.15%	15	45.45%	1	100%	34	46.58%
	same ethnicity	13	33.33%	8	24.24%	1	100%	22	30.14%
	different ethnicity	1	2.56%	3	9.09%	0	0%	4	5.48%
	doesn't matter	25	64.10%	22	66.67%	0	0%	47	64.38%
Do you prefer the character to look like you?	Yes	11	28.21%	10	30.30%	0	0%	21	28.77%
	No	6	15.38%	7	21.21%	0	0%	13	17.81%
	Doesn't matter	22	56.41%	16	48.48%	1	100%	39	53.42%

4.2. Study Stress Score Result

At baseline before our intervention, study stress scores are similar (Group 1 mean=5.23, s.d=2.44; Group 2 mean=5.04, s.d=2.66; Group 3 mean=5.38, s.d=2;). Categorised scores (low, medium, high), see Table 6, reveal that most students are moderately to highly stressed about their studies.

Table 6. Study Stress Score Result by Group

		Group1		Group2		Group3		Total	
		N	%	N	%	N	%	N	%
Score1	Low	5	25%	7	29.17%	3	12.50%	15	22.06%
	Mid	7	35%	9	37.50%	12	50%	28	41.18%
	High	8	40%	8	33.33%	9	37.50%	25	36.76%
Score2	Low	5	27.78%	12	50%	3	12%	20	29.85%
	Mid	11	61.11%	7	29.17%	16	64%	34	50.75%
	High	2	11.11%	5	20.83%	6	24%	13	19.40%
Score3	Low	6	37.50%	16	66.67%	3	12%	25	38.64%
	Mid	8	50%	6	25%	17	68%	31	47.69%
	High	2	12.50%	2	8.33%	5	20%	9	13.85%

Table 6 and the corresponding Figure 3 show the changes in the user's study stress state (score) in each group over time: baseline (score 1), after scenario 1 (score 2), after scenario 2 (score 3). Comparing the three conditions shows that the control group (Group 3) has the least score changes after the second scenario. Only 66.67% of participants in Group 2 had low emotional feeling towards their study at the end of the experiment. However, 37.5% of Group 1 and 12% of the control group reported low study stress by the end of the experiment. Considering group1 in Table 6, 40% of students had high stress score before interaction with

Sarah (score 1) and this amount dropped down to 11.11% and 12.50% after first and second interactions respectively. In group 2 we can see a different pattern, where 33.33% of students had high stress score at score1, then it went down to 20.83% in score 2 and 8.33% in score 3. The high stress score 2 in group 2 is double score 2 in group 1. High stress score in group 3 had the least reduction among all groups.

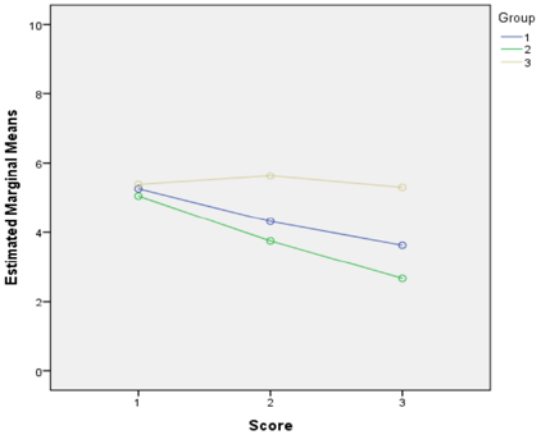

Figure 3: Study Stress Score Mean changes over the time

As can be seen in Figure 3, the change in stress levels is significantly larger for the agent treatments compared to the control. But since we used a crossover for the first two groups, we then separated the analysis into the two treatments (Empathic and Neutral). To compare the score mean differences between the groups we performed mixed Anova test. Figure 3 indicates that the variable group is an important factor. The within subject test indicates that each interaction has a significant effect on the study stress score, $F(2,122) = 20.6$, $p<0.005$, partial $\eta 2= 0.252$; in other words, the scores do change over time after experiencing scenario 1 and scenario 2. Moreover, the interaction of score and group is statistically significant $F(4,122) = 6.14$, $p <0.005$, partial $\eta 2= 0.168$) which means that the scores are changing over time depending on the group.

4.3. Relationship between Individual Factors

Cross tabs/contingency tables were used to analyse categorical data. A chi-square test revealed a significant relationship between the user's gender and the preference for the character's gender ($\chi^2(N=73)=20.11$, $p=0.00$). We found significant relationships between openness and stress ($\chi^2(N=73)=15.79$, $p=0.05$), emotional stability and depression ($\chi^2(N=73)=34.34$, $p=0.00$), emotional stability and anxiety ($\chi^2(N=73) = 20.32$, $p=0.01$) and emotional stability-stress ($\chi^2(N=73) = 24.47$, $p=0.00$) were identified using chi-square tests.

Regarding personality and study stress score, Chi-Square tests confirmed significant association between emotional stability and baseline study stress (score1), $\chi^2(N=68)=10.62$, $p=0.03$, Openness and study stress after scenario1 (score2), $\chi^2 (N=65)=9.48$, $p=0.05$ and emotional stability and score 2, $\chi^2 (N=67)=11.67$, $p=0.02$.

Cross tabulations between their stress score from the DASS21 and their initial level of study-related stress show that 60% of the students having extremely severe stress had medium level of initial study stress and 50% of the ones who had severe stress had high level of initial study stress. Group and study stress score after the first scenario (score2) were significantly related ($\chi^2(N=67)=10.34$, $p=0.04$). Moreover, group and study stress score after the second scenario (score3) were also significantly related ($\chi^2(N=65)=15.52$, $p=0.00$).

Participants with low level of openness reported high level of study stress at time2 (100%) and participants who had high level of openness reported low level of study stress (42%) at time2. Participants with low level of emotional stability reported high level of study stress at time1 (55%) and participants with high level of emotional stability reported low level of study stress at time1 (75%). Similarly, at time 2, participants with low level of emotional stability reported high level of study stress (40%) and participants with high level of emotional stability reported low level of study stress (75%). The only significant gender and personality differences were found for agreeableness, one of the personality dimension ($\chi^2(N=73)=16.27$, $p=0.00$).

At the end of the first scenario/session 1, participants were asked “Have I been able to help you?” (Groups 1 and 2) or “Was that helpful” (Group 3). Results are shown in Table 7. Analysis suggest that participant with higher openness and agreeableness were more likely to say they felt better. The numbers are too small in many categories for meaning chi-square test results.

Table 7. Have I been able to help you/Was that helpful?

	Yeah I feel better.	I feel about the same.	No, I feel worse now.
1- Empathic-Neutral	72%	28%	0%
2- Neutral-Empathic	83%	N/A	17%
3- Control	56%	40%	4%

In Group 1 and 2 at the end of the session, Sarah asked “Did you find our discussion useful?” and in the control group there was a question in the survey that asked “Did you find the document useful?” As shown in Table 8, in total 85% of the participants found the session useful. More participants found the conversation useful in group 1 & 2 than group 3 (87.5% and 87% vs 80%). Note that the question was asked at the conclusion of the experiment, after neutral interaction in group 1 and after empathic interaction in group 2. Analysis with individual factors, show that 80% of participants who did not find the document helpful had normal or moderate stress compared to 66% in these categories who found the document helpful, suggesting lower stress reduces the value of the study tips.

Table 8. Did you find our discussion/the document useful?

Group	Yeah it helped.	Not much.
1- Empathic-Neutral	87.5%	12.5%
2- Neutral-Empathic	87%	13%
3- Control	80%	20%
Total	85%	15%

5. Discussion

In general, our participants can be described as being medium (79.45%) to high (17.81%) open to new experiences/ideas, medium (79.45%) conscientiousness, spread on the extraversion dimension (ranging from introvert (21.92%) to extrovert (15.07), mostly moderate (63.01%)); low (27.4%) to moderate (67.12%) for emotional stability (or conversely high to moderate neuroticism). The sample population was more diverse concerning agreeableness and gender differences were found. Analysis of the data shows that 75.76% and 21.21% of males said they were moderately and highly agreeable, respectively, compared to 33.33% and 66.67% of females, respectively, no gender differences were found for other personality dimensions. This adds validity to our data, as personality is considered a fundamental individual trait, not restricted or connected to a particular gender or other individual factor. Gender differences for agreeableness may be due to genders perceiving themselves differently, possibly due to cultural gender biases regarding the social acceptability of being disagreeable varying for the genders. The literature supports our self-reported findings that females tend to be more cooperative and agreeable than males [36].

In answer to our first research question we did find some differences in preferences for and responses to the IVAs based on the participants’ gender, personality or psychological state. Our age distribution was too narrow to draw any conclusions regarding age-related preferences or responses. Below we discuss some of our findings and how they might be used to answer the second research question concerning modeling the user and tailoring IVA behaviour to respond to the user model. In general, our results indicate that there is no preference for the character’s ethnicity or similarity to the participants. Most of the participants (73.97%) preferred a peer-aged character. Almost half of both genders, preferred an IVA of the same gender and the other half did not care about gender. However, in the literature we most commonly find the use of female IVA models because the literature reports that in line with findings that female physicians are associated with empathic communication and relationship building [37-39]. The

implication of our results relating to our second research question is that tailoring of models according to needs and preferences, including gender matched models if desired, should be provided in deployed applications.

Differences were found in preferences for characters according to the individual's personality and emotional psychological state. For example, high levels of stress (reported via DASS21) showed significant differences regarding ethnicity preferences, for example, in the extremely severe stress category (7% of the total sample) 0% did not care about the ethnicity and 80% and 20% preferred a same and a different ethnicity respectively. In contrast, 76%,24%,0% with normal stress, 45%,36%,19% with mild stress, 43%,43%,14% with moderate stress and 89%,11%,0% with severe stress, did not care about ethnicity, preferred the same ethnicity or preferred a different ethnicity, respectively. There were no differences in participants' result of DASS21 and their cultural group. Moreover, we couldn't find any relation between participants' ethnicity and their ethnicity preferences for IVA. However, we found differences concerning study stress levels relating to personality. For example, significantly more individuals who were emotionally unstable (i.e. they were medium or high on the neurotic scale) reported feeling high stress (55% and 32%) and medium stress (35% and 44%, respectively) compared to emotionally stable (10% and 23%). However, the participants who were low on the neurotic scale reported less stress (75%).

Some personal factors did not reveal any significant relationships. A Chi-square test reported no significant relationship between the participant playing computer games and the study stress score. We did not find any significant differences between participants' age, gender and study stress score. Finally, a Chi-square test reported no significant relationship between the ethnicity of the agent and study stress score.

We found a relationship between the psychological state of the user (DASS21) and the study stress score reported by the user. The distribution of participants across the depression dimension was equal. The individuals with extremely severe scale of depression reported medium level of study stress (92%) at the end of the experiment while the participants with normal scale of depression reported low level of study stress (63%). Regarding stress dimension (reported via DASS21), most participants with severe and extremely severe scale of the stress reported medium level of study stress score at the end of the experiment (86% and 80%) while participants with normal scale of stress reported low study stress (57%). The test also shows that participants with low emotional stability (high neuroticism) are more likely to have high and severe depression, anxiety and stress. Moreover, participants with low openness are more likely to have severe stress. Our finding is inconsistent with the literature [40] where personality traits have been examined as predictors of depression, anxiety and stress.

The key goal of our agent in this experiment was to "Reduce Study Stress" and the IVAs were shown to achieve that more successfully than reading a pdf file with the same tips. Our results (Table 8) also show that participants found the both empathic and neutral conversation more useful than the pdf file. The results for both agent groups are similar and since this question is asked after having received both the empathic and neutral dialogues, our results are inconclusive regarding whether empathic dialogue was more useful than the neutral dialogue. The question regarding if they felt better after the first round (Table 7) indicates that all participants who received the empathic agent the same or better, whereas 17% with the neutral agent felt worse, compared to 4% who received the document.

In answer to our second research question relating to possible IVA adaptations to individual difference, these findings suggest that IVAs dealing with emotionally unstable, or more neurotic individuals, may need to take into account possibly higher levels of stress and exhibit more empathic or other stress reducing behaviours. An adaptive IVA may need to show more empathic behaviours for depressed and stressed individuals to minimise their stress level. In the current study we randomly assigned individuals to groups. We intend to use datamining methods on the dataset from this study (and another dataset) to discover IVA preference rules based on participant features and whether the document, neutral or empathic delivery of study tips was most useful for certain combination of individual features. For instance, preference rules could be like the following rules:

"Rule1: If 20<age<25 AND study_stress=medium AND emotional_state=high_anxiety THEN character_style=Empathic;

Rule2: If personality=introvert AND ethnicity=Oceania AND study_stress=low THEN character_style=Neutral;"

In our next study, we intend to use these rules to allocate participants to treatment groups (i.e. alternative IVAs). For example, we may assign people to an agent with different types of dialogue (neutral, empathic, highly empathic) according to their personality, the intensity of their emotions and stress levels and, if they have a preference, also provide participants with a male or female character according to their preference.

Clearly, there are many factors that come into play when evaluating the relation between an IVA and the human user. According to [17] students prefer to interact with an APA with more social cues (i.e. facial expression, vocal tone and body language) rather than an APA with only auditory capability. In this study, although our virtual character did not display many facial expressions but she expressed emotions and built rapport with her human-like representation. The FAtiMA model allows researchers to develop rules of user interaction in relation to the agent's emotional state (which can be empathetic or otherwise), the state of the environment they are in and various social norms, values and rituals that can be designed to mimic various cultural protocols and human-human interaction styles [35,3]. The aim is for agents to create trust; put the user at ease; and convince them of the agent's ability to guide. The research reported here is a part of initiatives to evaluate agent-human interaction. The use of agent models will improve the agent's ability to change to fit the demands of the situation they are in. However, it will require more work to develop the rules that are suitable for different user scenarios.

While we have found some significant results and suggested some tailoring, our study has a number of limitations requiring further future investigation and studies. The majority of our participants are psychology students (54/73) and, related to the first limitation, they are predominantly female, though the gender distribution overall is nearly balanced (39 female, 33 male). Furthermore, we need more participants to represent more age groups, cultures and more equal distributions of experimental groups.

6. Conclusion and Future Work

This study, together with other studies reported in the literature, strongly suggest that IVAs need to adapt in appearance and behaviour according to their purpose. In some cases, the IVA needs to match the user's preferences (i.e. gender, age, culture, similarity to the user), but in other cases where behaviour change or biases are to be challenged, then the IVA should become whatever is most beneficial in achieving that goal (e.g. empathic, neutral, etc.).

To design IVAs that become what the human needs them to be, we require much more sophisticated models of the human as part of our agent architectures. The human should not be seen as a component of the environment and source of input, but be modeled within the architecture and have reasoning modules and rules that enable the IVA to behave accordingly. With increasing access to user data and the growth in the Internet of Things and ambient computing, capturing data about our user's preferences, beliefs, emotional state, exercise regime and diet, etc. should become possible. Now is the time to conduct studies to understand what data is important and how it can best to used. This study seeks to make such a contribution.

References

1. DeVault, D., et al., *SimSensei kiosk: a virtual human interviewer for healthcare decision support*, in *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*2014, International Foundation for Autonomous Agents and Multiagent Systems: Paris, France. p. 1061-1068.

2. Cerekovic, A., O. Aran, and D. Gatica-Perez, *Rapport with virtual agents: What do human social cues and personality explain?* IEEE Transactions on Affective Computing, 2016.
3. Lim, M.Y., et al. *Improving adaptiveness in autonomous characters*. in *International Workshop on Intelligent Virtual Agents*. 2008. Springer.
4. Kutay, C., et al. *Intercultural-role Plays for e-Learning using Emotive Agents*. in *ICAART (2)*. 2013.
5. Lisetti, C.L., *10 advantages of using avatars in patient-centered computer-based interventions for behavior change.* SIGHIT Record, 2012. **2**(1): p. 28.
6. Shen, L., M. Wang, and R. Shen, *Affective e-learning: Using" emotional" data to improve learning in pervasive learning environment.* Journal of Educational Technology & Society, 2009. **12**(2): p. 176.
7. Johnson, W.L., J.W. Rickel, and J.C. Lester, *Animated pedagogical agents: Face-to-face interaction in interactive learning environments*. International Journal of Artificial intelligence in education, 2000. **11**(1): p. 47-78.
8. Lester, J.C., et al. *The persona effect: affective impact of animated pedagogical agents*. in *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*. 1997. ACM.
9. Veletsianos, G., *Cognitive and affective benefits of an animated pedagogical agent: Considering contextual relevance and aesthetics.* Journal of Educational Computing Research, 2007. **36**(4): p. 373-377.
10. Woolf, B., et al., *Affect-aware tutors: recognising and responding to student affect.* International Journal of Learning Technology, 2009. **4**(3-4): p. 129-164.
11. Sabourin, J., B. Mott, and J. Lester. *Computational models of affect and empathy for pedagogical virtual agents*. in *Standards in emotion modeling, Lorentz Center International Center for workshops in the Sciences*. 2011.
12. Kim, Y. *Empathetic virtual peers enhanced learner interest and self-efficacy*. in *Workshop on Motivation and Affect in Educational Software, in conjunction with the 12th International Conference on Artificial Intelligence in Education*. 2005.
13. Prendinger, H. and M. Ishizuka, *The empathic companion: A character-based interface that addresses users'affective states.* Applied Artificial Intelligence, 2005. **19**(3-4): p. 267-285.
14. D'mello, S.K. and A. Graesser, *Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features.* User Modeling and User-Adapted Interaction, 2010. **20**(2): p. 147-187.
15. Ammar, M.B., et al., *The affective tutoring system.* Expert Systems with Applications, 2010. **37**(4): p. 3013-3023.
16. Duo, S. and L.X. Song, *An e-learning system based on affective computing.* Physics Procedia, 2012. **24**: p. 1893-1898.
17. Moreno, R. and T. Flowerday, *Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity.* Contemporary educational psychology, 2006. **31**(2): p. 186-207.
18. dos Santos, C.T. and F.S. Osório. *An intelligent and adaptive virtual environment and its application in distance learning*. in *Proceedings of the working conference on Advanced visual interfaces*. 2004. ACM.
19. Caridakis, G., et al., *Virtual agent multimodal mimicry of humans.* Language Resources and Evaluation, 2007. **41**(3-4): p. 367-388.
20. Bevacqua, E., M. Mancini, and C. Pelachaud. *A listening agent exhibiting variable behaviour*. in *Intelligent Virtual Agents*. 2008. Springer.
21. Pecune, F., et al. *Laughing with a Virtual Agent*. in *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. 2015. International Foundation for Autonomous Agents and Multiagent Systems.
22. Kistler, F., et al., *Natural interaction with culturally adaptive virtual characters.* Journal on Multimodal User Interfaces, 2012. **6**(1-2): p. 39-47.

23. Bailenson, J.N., et al., *The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments.* Presence: Teleoperators and Virtual Environments, 2005. **14**(4): p. 379-393.
24. Vugt, H.C.V., et al., *Effects of facial similarity on user responses to embodied agents.* ACM Transactions on Computer-Human Interaction (TOCHI), 2010. **17**(2): p. 7.
25. Bickmore, T.W., D. Schulman, and C. Sidner, *Automated interventions for multiple health behaviors using conversational agents.* Patient education and counseling, 2013. **92**(2): p. 142-148.
26. Lisetti, C., et al., *I can help you change! an empathic virtual agent delivers behavior change health interventions.* ACM Transactions on Management Information Systems (TMIS), 2013. **4**(4): p. 19.
27. Jin, S.-A.A., *The effects of incorporating a virtual agent in a computer-aided test designed for stress management education: The mediating role of enjoyment.* Computers in Human Behavior, 2010. **26**(3): p. 443-451.
28. Ducheneaut, N., et al. *Body and mind: a study of avatar personalization in three virtual worlds*. in *Proceedings of the SIGCHI conference on human factors in computing systems*. 2009. ACM.
29. Hall, L., et al., *Designing empathic agents: Adults versus kids.* Intelligent tutoring systems, 2004: p. 604-613.
30. Rivera-Gutierrez, D., et al., *Getting the Point Across: Exploring the Effects of Dynamic Virtual Humans in an Interactive Museum Exhibit on User Perceptions.* IEEE transactions on visualization and computer graphics, 2014. **20**(4): p. 636-643.
31. Dias, J., S. Mascarenhas, and A. Paiva, *Fatima modular: Towards an agent architecture with a generic appraisal framework*, in *Emotion Modeling*. 2014, Springer. p. 44-56.
32. Mancini, M., et al. *Greta: a SAIBA compliant ECA system*. in *Troisiéme Workshop sur les Agents Conversationnels Animés*. 2008.
33. Bickmore, T., A. Gruber, and R. Picard, *Establishing the computer–patient working alliance in automated health behavior change interventions.* Patient education and counseling, 2005. **59**(1): p. 21-30.
34. Goldberg, L.R., *A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models.* Personality psychology in Europe, 1999. **7**(1): p. 7-28.
35. Henry, J.D. and J.R. Crawford, *The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample.* British journal of clinical psychology, 2005. **44**(2): p. 227-239.
36. Weisberg, Y.J., C.G. DeYoung, and J.B. Hirsh, *Gender differences in personality across the ten aspects of the Big Five.* Frontiers in psychology, 2011. **2**: p. 178.
37. Baker, S., D. Richards, and P. Caldwell, *Relational Agents to Promote eHealth Advice Adherence*, in *PRICAI 2014: Trends in Artificial Intelligence*. 2014, Springer. p. 1010-1015.
38. Baker, S., D. Richards, and P. Caldwell, *Putting a New Intelligent Virtual Face on a Medical Treatment Advice System to Improve Adherence*. Proceedings of the 2014 Conference on Interactive Entertainment, 2014: p. 1-9.
39. Schmid Mast, M., J.A. Hall, and D.L. Roter, *Disentangling physician sex and physician communication style: Their effects on patient satisfaction in a virtual medical visit.* Patient Education and Counseling, 2007. **68**(1): p. 16-22.
40. Vujičić, M.M. and D.J. Ranđelović, *Personality traits as predictors of depression, anxiety, and stress with secondary school students of final years.* Zbornik radova Filozofskog fakulteta u Prištini, 2017(47-3): p. 217-237.